

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re U.S. Patent Application of )  
TAGO et al. )  
Application Number: To be Assigned )  
Filed: Concurrently Herewith )  
For: KEY WORD FREQUENCY CALCULATION METHOD )  
AND PROGRAM FOR CARRYING OUT THE SAME )  
ATTORNEY DOCKET NO. HIRA.0143 )

Honorable Assistant Commissioner  
for Patents  
Washington, D.C. 20231

**REQUEST FOR PRIORITY  
UNDER 35 U.S.C. § 119  
AND THE INTERNATIONAL CONVENTION**

Sir:

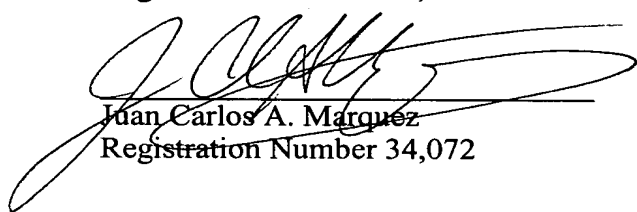
In the matter of the above-captioned application for a United States patent, notice is hereby given that the Applicant claims the priority date of March 28, 2003, the filing date of the corresponding Japanese patent application 2003-092098.

A certified copy of Japanese patent application 2003-092098 is being submitted herewith. Acknowledgment of receipt of the certified copy is respectfully requested in due course.

Respectfully submitted,

\_\_\_\_\_  
Stanley P. Fisher  
Registration Number 24,344

**REED SMITH LLP**  
3110 Fairview Park Drive  
Suite 1400  
Falls Church, Virginia 22042  
(703) 641-4200  
February 11, 2004

  
\_\_\_\_\_  
Juan Carlos A. Marquez  
Registration Number 34,072

(Translation)

**PATENT OFFICE  
JAPANESE GOVERNMENT**

**This is to certify that the annexed is a true copy of the following application as  
filed with this Office.**

**Date of Application:        March 28, 2003**

**Application Number:        Japanese Patent Application  
                                     No. 2003-92098**

**Applicant(s):                Hitachi Software Engineering Co., Ltd.**

**October 22, 2003**

**Commissioner,  
Patent Office**

**Yasuo IMAI (seal)**

**Certificate No.2003-3087348**

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日            2 0 0 3 年    3 月 2 8 日  
Date of Application:

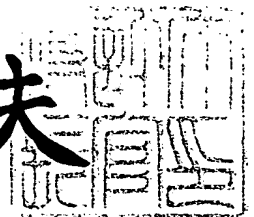
出 願 番 号            特 願 2 0 0 3 - 0 9 2 0 9 8  
Application Number:  
[ST. 10/C] :            [ J P 2 0 0 3 - 0 9 2 0 9 8 ]

出 願 人            日 立 ソ フ ト ウ エ ア エ ン ジ ニ ア リ ン グ 株 式 有 限 公 司  
Applicant(s):

2 0 0 3 年 1 0 月 2 2 日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

今 井 康 夫



出 証 番 号    出 証 特 2 0 0 3 - 3 0 8 7 3 4 8

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年 3月28日  
Date of Application:

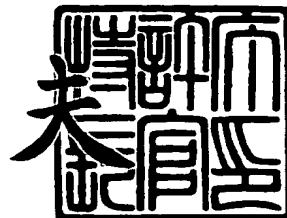
出願番号 特願2003-092098  
Application Number:  
[ST. 10/C]: [JP2003-092098]

出願人 日立ソフトウェアエンジニアリング株式会社  
Applicant(s):

2003年10月22日

特許庁長官  
Commissioner,  
Japan Patent Office

今井 康



【書類名】 特許願

【整理番号】 15A003

【提出日】 平成15年 3月28日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/30

【発明の名称】 キーワード頻度算出方法及びそれを実行するプログラム

【請求項の数】 4

【発明者】

    【住所又は居所】 東京都品川区東品川4丁目12番7号 日立ソフトウェアエンジニアリング株式会社内

    【氏名】 多胡 滋

【発明者】

    【住所又は居所】 東京都品川区東品川4丁目12番7号 日立ソフトウェアエンジニアリング株式会社内

    【氏名】 吉井 淳治

【発明者】

    【住所又は居所】 東京都品川区東品川4丁目12番7号 日立ソフトウェアエンジニアリング株式会社内

    【氏名】 水沼 貞

【特許出願人】

    【識別番号】 000233055

    【氏名又は名称】 日立ソフトウェアエンジニアリング株式会社

【代理人】

    【識別番号】 100091096

    【弁理士】

    【氏名又は名称】 平木 祐輔

【選任した代理人】

【識別番号】 100102576

【弁理士】

【氏名又は名称】 渡辺 敏章

【選任した代理人】

【識別番号】 100100169

【弁理士】

【氏名又は名称】 大塩 剛

【手数料の表示】

【予納台帳番号】 015244

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9722155

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 キーワード頻度算出方法及びそれを実行するプログラム

【特許請求の範囲】

【請求項 1】 塩基配列及びアミノ酸配列に関する情報を格納した第 1 のデータベースと文書データを格納した第 2 のデータベースとを用いてキーワードの出現頻度を算出するキーワード頻度算出方法であって、ユーザが入力した塩基配列又はアミノ酸配列に基づいて、第 1 のデータベースより第 1 のテキストデータを抽出する第 1 のテキストデータ抽出ステップと、上記抽出した第 1 のテキストデータよりそれに含まれる文書データを識別する識別子を抽出する識別子抽出ステップと、上記抽出した識別子に基づいて、第 2 のデータベースより第 2 のテキストデータを抽出する第 2 のテキストデータ抽出ステップと、上記第 1 のデータベースに関連したキーワードを含むキーワードテーブルよりキーワードを順次読み出し、上記第 2 のテキストデータにおける上記各キーワードの出現頻度を算出する出現頻度算出ステップと、を含むキーワード頻度算出方法。

【請求項 2】 請求項 1 記載のキーワード頻度算出方法において、上記キーワードテーブルは、キーワードをカテゴリー毎に分類して格納したツリー構造を有し、上記出現頻度算出ステップは、キーワードの出現頻度とキーワードが属する上位のカテゴリーの出現頻度を含むツリー構造の頻度算出結果テーブルを生成することを特徴とするキーワード頻度算出方法。

【請求項 3】 第 1 のテキストデータ抽出ステップは、ユーザが複数の配列を入力したとき、各配列に関して第 1 のデータベースより第 1 のテキストデータを抽出することを特徴とするキーワード頻度算出方法。


【請求項 4】 請求項 1 から 3 のいずれか 1 項記載のキーワード頻度算出方法をコンピュータに実行させるためのプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は遺伝子関連のデータの検索に利用して好適なデータベース検索技術に関し、特に、テキストマイニング方法によって文書データに含まれるキーワード



の頻度を検出するためのデータベース検索技術に関する。

**【0002】****【従来の技術】**

遺伝子またはタンパク質に関する研究成果を記述した文書データのデータベースには一般に2種類が存在する。第1のデータベースには研究のテーマとしている塩基配列またはアミノ酸配列が記述されており、第2のデータベースにはその配列を持つ遺伝子またはタンパク質に関する機能や性質が記述されている。そして、第1のデータベースのデータには塩基配列情報またはアミノ酸配列情報とともに、同一の遺伝子またはタンパク質に関して記述している第2のデータベース上の文書データの識別子が関連文書データとして記述されているのが一般的である。

**【0003】**

一方、ある特定の遺伝子またはタンパク質の機能または性質を知りたい検索者に対し、次のいずれかの方法がこれまで提供されている。第1の方法では、その遺伝子またはタンパク質の配列情報を検索キーとして、前記第1のデータベースを検索し、得られた第1のデータベースのデータ中から第2のデータベースのデータの識別子を抽出し、第2のデータベースのデータを取得し、それを検索者が参照してそこに記述されている各遺伝子またはタンパク質の機能または性質を理解する。この方法の例として「BLAST」(<http://www.ncbi.nlm.nih.gov/BLAST/>)と呼ばれる方法が広く利用されている。

**【0004】**

第2の方法では、特定の遺伝子またはタンパク質の識別子またはそれに類する関連情報を配列情報とは異なるキーワードとして1つまたは複数選択し、そのキーワードのいずれかを含む第2のデータベースのデータを抽出し、それを検索者が参照してそこに記述されている各遺伝子またはタンパク質の機能または性質を理解する。さらに、抽出された第2のデータベースのデータを少数に絞り込むための方法として知識に相当する情報を適用する方法の例として、「特開2002-32374 情報抽出方法及び記録媒体」がある。

**【特許文献1】**



特開 2002-32374 号公報

## 【0005】

## 【発明が解決しようとする課題】

しかしながら、従来の方法には次のような課題がある。すなわち、第1の方法では、第2のデータベースのデータの参照を検索者が直接行わなければならないため、特定の遺伝子またはタンパク質の機能または性質を把握するために大量の文書データを参照しなければならない。

## 【0006】

また、第2の方法では、適切なキーワードが選択可能であれば、適切な文書データグループを抽出することが可能となるものの、特定の塩基配列またはアミノ酸配列を持つ遺伝子またはタンパク質がいかなる機能または性質を持つのか知識のない検索者にとっては適切なキーワードを選択することは困難である。逆に言えば特定の遺伝子またはタンパク質がいかなる機能または性質を持つのかを知りたい検索者が検索を実行するのであり、やはりキーワードを検索者自身が選択することは困難である。従って、適切な文書データグループを抽出することは難しい。

## 【0007】

## 【課題を解決するための手段】

本発明によると、塩基配列及びアミノ酸配列に関する情報を格納した第1のデータベースと文書データを格納した第2のデータベースとを用いてキーワードの出現頻度を算出するキーワード頻度算出方法であって、ユーザが入力した塩基配列又はアミノ酸配列に基づいて、第1のデータベースより第1のテキストデータを抽出する第1のテキストデータ抽出ステップと、上記抽出した第1のテキストデータよりそれに含まれる文書データを識別する識別子を抽出する識別子抽出ステップと、上記抽出した識別子に基づいて、第2のデータベースより第2のテキストデータを抽出する第2のテキストデータ抽出ステップと、上記第1のデータベースに関連したキーワードを含むキーワードテーブルよりキーワードを順次読み出し、上記第2のテキストデータにおける上記各キーワードの出現頻度を算出する出現頻度算出ステップと、を含む。

**【0008】**

本発明によると、検索者は特定の配列を持つ遺伝子またはタンパク質について、その機能または性質を知りたい場合に、配列情報自体を検索キーとして入力することで、その遺伝子またはタンパク質の機能または性質を表すキーワードの一覧を、そのキーワードの文書データ中での出現頻度すなわち重要度として入手することが可能となる。

**【0009】**

また、複数の配列を検索キーとして入力することで、複数の遺伝子またはタンパク質に共通の機能または性質についても同様にそれを表すキーワードの一覧を入手することが可能となる。

**【0010】****【発明の実施の形態】**

以下図を参照して、本発明の実施例を説明する。図1は本発明によるデータベース検索システム構成を示す。本例のデータベース検索システムは、ディスプレイ装置101、演算装置102、マウス装置103、キーボード104、及び、第1、第2及び第3のファイルシステム105、107、109を有する。

**【0011】**

ディスプレイ装置101は、文字および図形を表示する機能とマウスカーソルを表示する機能とを有する。演算装置102は、ディスプレイ装置101上のマウスカーソルの位置を受信する機能と、キーボードから任意の文字列を受信する機能と、メモリ中にデータを保持する機能と、テキストデータから特定の一部の文字列を切り出す機能と、特定の文字列同士が一致するか否かを判定する機能を有する。マウス装置103は、ディスプレイ装置101上のマウスカーソルの移動を指示し、ボタンを押すことによりマウスカーソルの位置の認識を指示する機能を有する。キーボード104は、任意の文字列を入力し、演算装置102に送信する機能を有する。

**【0012】**

第1のファイルシステム105は、テキストデータ106を各々ファイルとして保持する機能を持つ補助記憶装置である。第2のファイルシステム107は、

テキストデータ 108 を各々ファイルとして保持する機能を持つ補助記憶装置である。第3のファイルシステム 109 は、カテゴリーテーブル 110 をファイルとして保持する機能を持つ補助記憶装置である。

#### 【0013】

図2は第1のファイルシステム 105 のテキストデータ 106 の構造を示している。この例では特定の塩基配列に関連する研究成果をまとめた論文状のデータを示している。テキストデータ 106 は、本データが記述対象としている塩基又はアミノ酸の配列 201 と、本データと関連のある記述がされている他のテキストデータの識別子 202 とを含む。この例では本データに対して2つの関連するテキストデータがあるので、2つの識別子が格納されている。この例では「PMID (PubMed ID)」と呼称する識別子を示す。

#### 【0014】

図3は第2のファイルシステム 107 のテキストデータ 108 の構造を示している。テキストデータ 108 は、本データの識別子 301 と、本データの本文に相当する文字列 302 とを含む。この例では、遺伝子またはタンパク質等、分子生物学上の研究成果について記述されているものとする。

#### 【0015】

図4はディスプレイ装置 101 上に表示される検索開始画面の構成を示している。検索開始画面は、ユーザが、塩基またはアミノ酸の配列を文字列として入力するフィールド 401 と、演算処理装置 102 に検索開始を指示するための検索開始ボタン 402 とを含む。

#### 【0016】

図5は第3のファイルシステム 109 のカテゴリーテーブル 110 の構造を示している。カテゴリーテーブル 110 は、1つまたは複数のキーワードが属するカテゴリーの名前を格納するカテゴリー部 501 とその下位のカテゴリーの名前を格納する下位カテゴリー部 502 とキーワードを格納するキーワード部 503 とを含む。カテゴリーテーブル 110 に含まれるキーワードは、第2のファイルシステム 107 のテキストデータ 108 に含まれる情報に関連したキーワードのみを含むものであってよい。この例では下位カテゴリー「axon guidance」及び

「axon extension」は上位のカテゴリ「cell recognition」に属することを示している。キーワード「motor axon guidance」は下位カテゴリ「axon guidance」に属することを示している。

#### 【0017】

再び図1を参照して、本発明によるデータベース検索システムの概念を説明する。先ずユーザは、キーボード104を使用して、塩基配列又はアミノ酸配列を入力する。例えば、塩基配列AGCTを入力する。演算装置102は、配列AGCTに基づいて、第1のファイルシステム105より、配列AGCTを含む又は配列AGCTに関連した情報を含むテキストデータ106を抽出する。

#### 【0018】

各テキストデータ106は、文書データを識別する識別子202を含む。演算装置102は、各テキストデータ106より、識別子202を抽出し、それを使用して、第2のファイルシステム107より、識別子202に対応したテキストデータ108を抽出する。

#### 【0019】

次に、演算装置102は、第3のファイルシステム109のカテゴリテーブル110に含まれるキーワードを取り出し、抽出したテキストデータ108におけるキーワードの出現頻度を算出する。即ち、抽出したテキストデータ108のうち、各キーワードが出現する又は使用されているテキストデータ108の数を算出する。

#### 【0020】

こうしてユーザは、第2のファイルシステム107のテキストデータ108において、配列AGCTに関連するキーワードの頻度を得ることができる。カテゴリテーブル110は、キーワードをカテゴリに分類したツリー構造にて格納している。従って、ユーザはディスプレイ装置101の画面にて、ツリー構造のキーワード頻度算出結果テーブルを得ることができる。

#### 【0021】

図6は、テキストデータ108において図5のキーワードに対する頻度算出結果テーブルを示す。図5と図6を比較すると明らかなように、頻度算出結果テ

ブルの領域 601 には、カテゴリーテーブル 110 のカテゴリー部 501 の各カテゴリーの頻度が表示され、領域 602 には、カテゴリーテーブル 110 の下位カテゴリー部 502 の各下位カテゴリーの頻度が表示され、領域 603 には、カテゴリーテーブル 110 のキーワード部 503 の各キーワードの頻度が表示されている。

#### 【0022】

カテゴリー部 501 の各カテゴリーの頻度は、それに属する下位カテゴリーの頻度の合計であり、下位カテゴリー部 502 の各下位カテゴリーの頻度は、それに属するキーワードの頻度の合計である。従って、領域 603 のキーワードの頻度が求められれば、それより上位の全てのカテゴリーの頻度は得られる。

#### 【0023】

この例では、「cell recognition」というカテゴリーに属する全キーワードの出現頻度は、196 である。これは、第 2 のファイルシステム 107 に含まれるテキストデータのうち、196 個のテキストデータにおいて、「cell recognition」というカテゴリーに属するキーワードが少なくとも 1 回以上出現していることを示している。

#### 【0024】

「motor axon guidance」というキーワードの出現頻度は 18 である。これは、第 2 のファイルシステム 107 に含まれるテキストデータのうち、「motor axon guidance」というキーワードを 1 回以上含むテキストデータの総数が 18 個であることを示している。

#### 【0025】

図 7 は、ディスプレイ装置 101 の画面に表示されているツリー構造のカテゴリー及びキーワードの頻度算出結果テーブルである。図 5 のカテゴリーテーブル 110 に、図 6 の頻度算出結果テーブルを重ね合わせることによって生成される。図 7 のツリー構造の頻度テーブルの領域 701 及び 702 は図 5 におけるカテゴリー 501 及び下位カテゴリー 502 に対応している図形ノードである。領域 703 は図 5 におけるキーワード 503 に対応している図形ノードである。

#### 【0026】

図 8 を参照して本発明によるデータベース検索方法の処理の流れを説明する。まず、ステップ 801 において、ユーザは、図 4 の検索開始画面の入力フィールド 401 に、塩基またはアミノ酸の配列を表現する文字列を入力する。図 4 の例では「A」「G」「C」「T」という 4 種類の塩基を一行に並べることでその配列を表現する形式となっている。ここで、複数の配列を入力する場合は、配列を表現する文字列の間に空行を挿入する。ユーザは、マウス装置 103 によって、図 4 の検索開始画面の検索開始ボタン 402 をクリックすることにより、次のステップ 802 に進む。

#### 【0027】

次に、ステップ 802 において、図 4 の検索開始画面の入力フィールド 401 にて入力された全ての配列が処理されたかどうかをチェックし、全ての配列が処理された場合にはステップ 814 に進み、全ての配列が処理されていない場合にはステップ 803 に進む。

#### 【0028】

次に、ステップ 803 において、第 1 のファイルシステム 105 からテキストデータファイル 106 を 1 つ取り出す。ステップ 804 において、全てのテキストデータファイルについて処理されたか否かを判定する。全てのテキストデータファイルについて処理された場合には、ステップ 802 に戻り、次の配列の処理を行う。全てのテキストデータファイルについて処理されていない場合には、ステップ 805 に進む。以下、ステップ 804 において全てのテキストデータファイルについて処理されたと判断されるまで、ステップ 803 以降の処理を繰り返す。

#### 【0029】

ステップ 805 において、ステップ 803 にて取り出したテキストデータファイル 106 より配列文字列 201 を取り出し、その配列文字列がステップ 801 で入力された配列文字列のうち現在処理の対象としている文字列と一致するかまたはその一部を含むか否かを調べる。調べる方法としては、前述の「BLAST」を使用してよい。配列文字列が含まれている場合には、ステップ 806 に進み、配列文字列が含まれていない場合には、ステップ 803 に戻り、次のファイルを取

り出して、以降のステップを処理する。

#### 【0030】

次に、ステップ806において、テキストデータファイル106より識別子202を取り出す。次に、ステップ807において、第2のファイルシステム107からテキストデータファイル108を1つ取り出す。ステップ808において、第2のファイルシステムの全てのテキストデータファイルについて処理が完了したか否かを判定する。第2のファイルシステムの全てのテキストデータファイルについて処理が完了した場合には、ステップ803に戻り、次のファイルを取り出し、上述の処理を行う。第2のファイルシステムの全てのテキストデータファイルについて処理が完了していない場合には、以下のステップを繰り返し実行する。

#### 【0031】

ステップ809において、テキストデータファイル107から本データの識別子301を取り出し、ステップ806にて取り出したテキストデータファイル107の識別子202のいずれかと一致するか否かを調べる。一致する場合にはステップ810に進み、一致していない場合には、ステップ807に戻り、別のファイルを取り出して処理を続行する。

#### 【0032】

ステップ810において、カテゴリーテーブル110からキーワードを1つ取り出す。そしてステップ811においてカテゴリーテーブルの全てのキーワードについて処理が完了したか否かを判定する。全てのキーワードについて処理が完了した場合はステップ807に戻り、他のファイルの処理を行う。全てのキーワードについて処理が完了していない場合はステップ812に進む。

#### 【0033】

次に、ステップ812において、ステップ807にて取り出したテキストデータファイルにステップ810で取り出したキーワードが含まれているか否かを調べる。含まれていない場合は、ステップ810に戻り、次のキーワードの処理を行い、含まれている場合は、ステップ813に進む。

#### 【0034】

ステップ 813 において、図 6 の頻度算出結果テーブルのキーワード出現頻度格納領域 603 における処理を行ったキーワードに対応した位置の頻度値を 1 増加させる。このとき、処理を行ったキーワードの上位カテゴリーに相当するカテゴリー 501 および 502 についても、それらに対応したキーワード出現頻度格納領域 601 および 602 の位置の頻度値を 1 増加させる。そしてステップ 810 に戻る。

このようにステップ 802 において全ての配列文字列について処理が完了したと判断された場合は、ステップ 814 に進む。

#### 【0035】

ステップ 814 において、図 5 のカテゴリーテーブルと図 6 の頻度算出結果テーブルの内容を反映させた図 7 のツリー構造の頻度テーブルをディスプレイ装置 101 上に表示させる。例えばマウス装置によっていずれかのカテゴリーに対応した図形ノードをクリックすると、その下位の図形ノードの表示および非表示を切り替えるなどの操作によって、使用者が適宜参照したい部分ツリーを表示させるようにする。

#### 【0036】

図 8 の処理はコンピュータによって実行されてよい。従って、本発明は図 8 の処理をコンピュータに実行させるためのプログラム及びそのようなプログラムを格納した記録媒体を含む。

#### 【0037】

以上本発明の例を説明したが本発明に上述の例に限定されるものではなく、特許請求の範囲に記載された発明の範囲にて様々な変更が可能であることは当業者に理解されよう。

#### 【0038】

##### 【発明の効果】

本発明によると、検索者は特定の配列を持つ遺伝子またはタンパク質について、その機能または性質を知りたい場合に、配列情報自体を検索キーとして入力することで、その遺伝子またはタンパク質の機能または性質を表すキーワードの一覧を、そのキーワードの文書データ中での出現頻度すなわち重要度として入手す



ることが可能となる効果がある。

### 【0039】

本発明によると、複数の配列を検索キーとして入力することで、複数の遺伝子またはタンパク質に共通の機能または性質についても同様にそれを表すキーワードの一覧を入手することが可能となる効果がある。

### 【図面の簡単な説明】

#### 【図1】

本発明によるデータベース検索システムの構成を示す図である。

#### 【図2】

第1のテキストデータファイルの構造を示す図である。

#### 【図3】

第2のテキストデータファイルの構造を示す図である。

#### 【図4】

配列文字列入力画面の例を示す図である。

#### 【図5】

カテゴリーテーブルの構造を示す図である。

#### 【図6】

頻度算出結果テーブルの構造を示す図である。

#### 【図7】

ツリー構造の頻度テーブルの構造を示す図である。

#### 【図8】

本発明によるデータベース検索システムの動作を示す流れ図である。

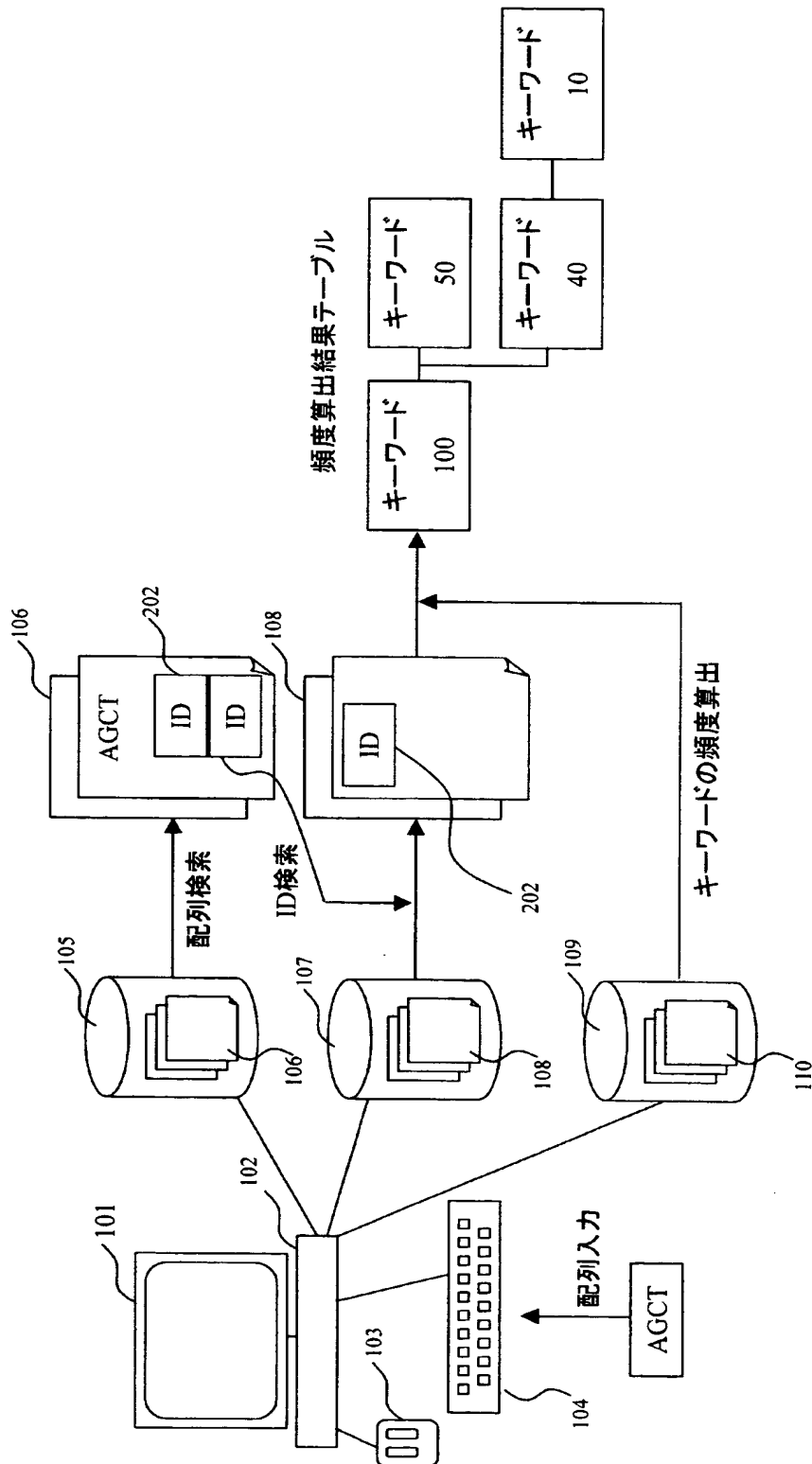
### 【符号の説明】

101…ディスプレイ装置、 102…演算装置、 103…マウス装置、 104…キーボード、 105, 107, 109…ファイルシステム 106, 108, 110…テキストデータファイル

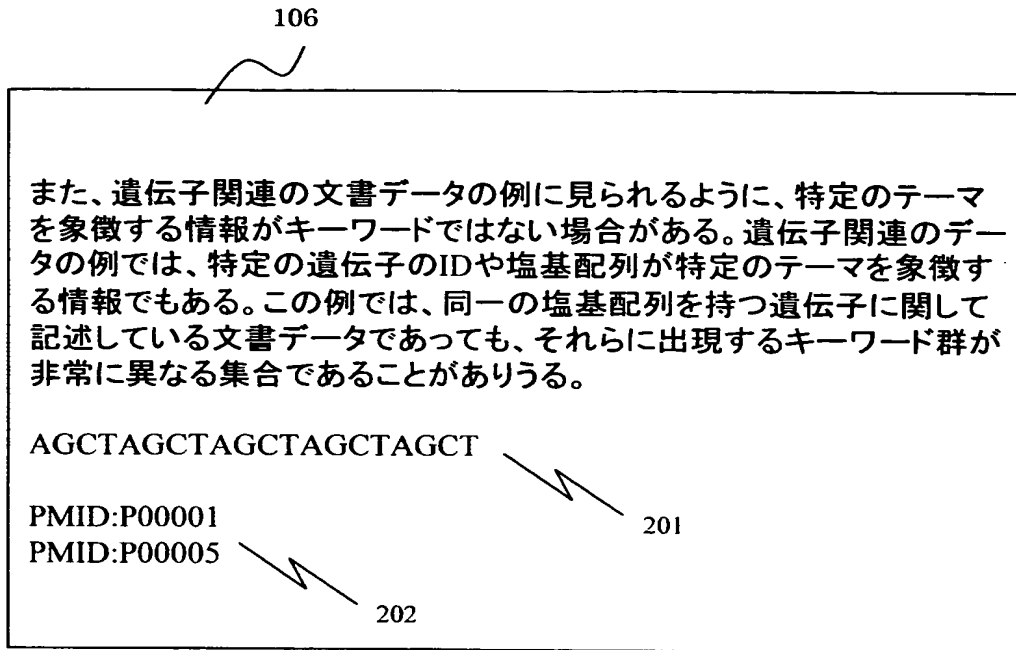
【書類名】

図面

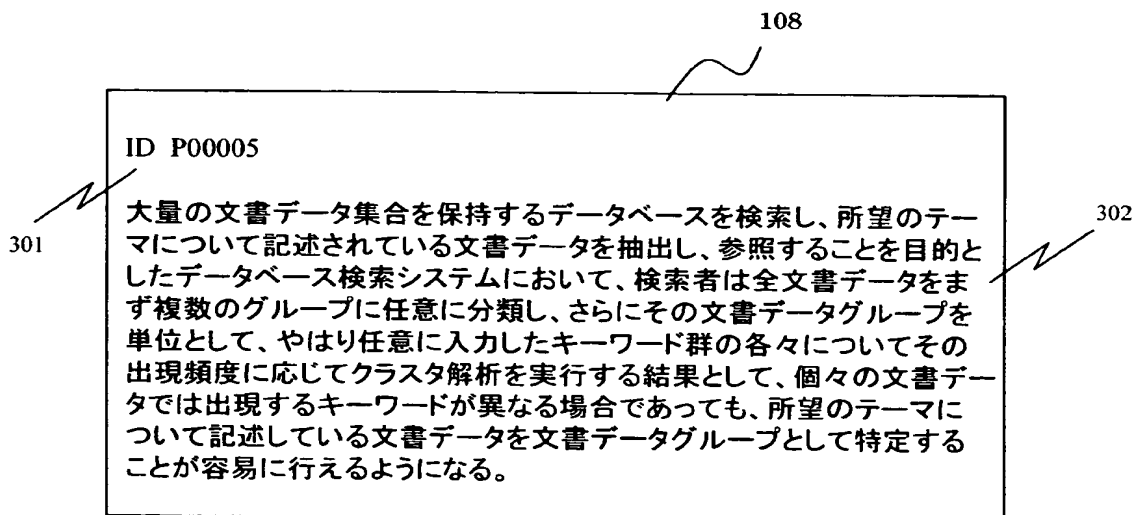
【図 1】



【図 2】



【図 3】

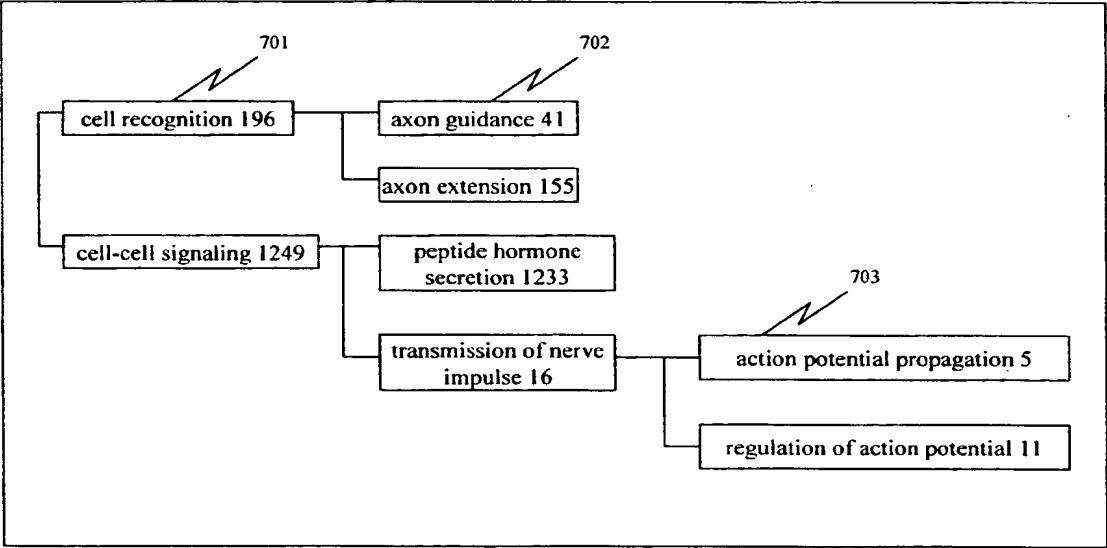




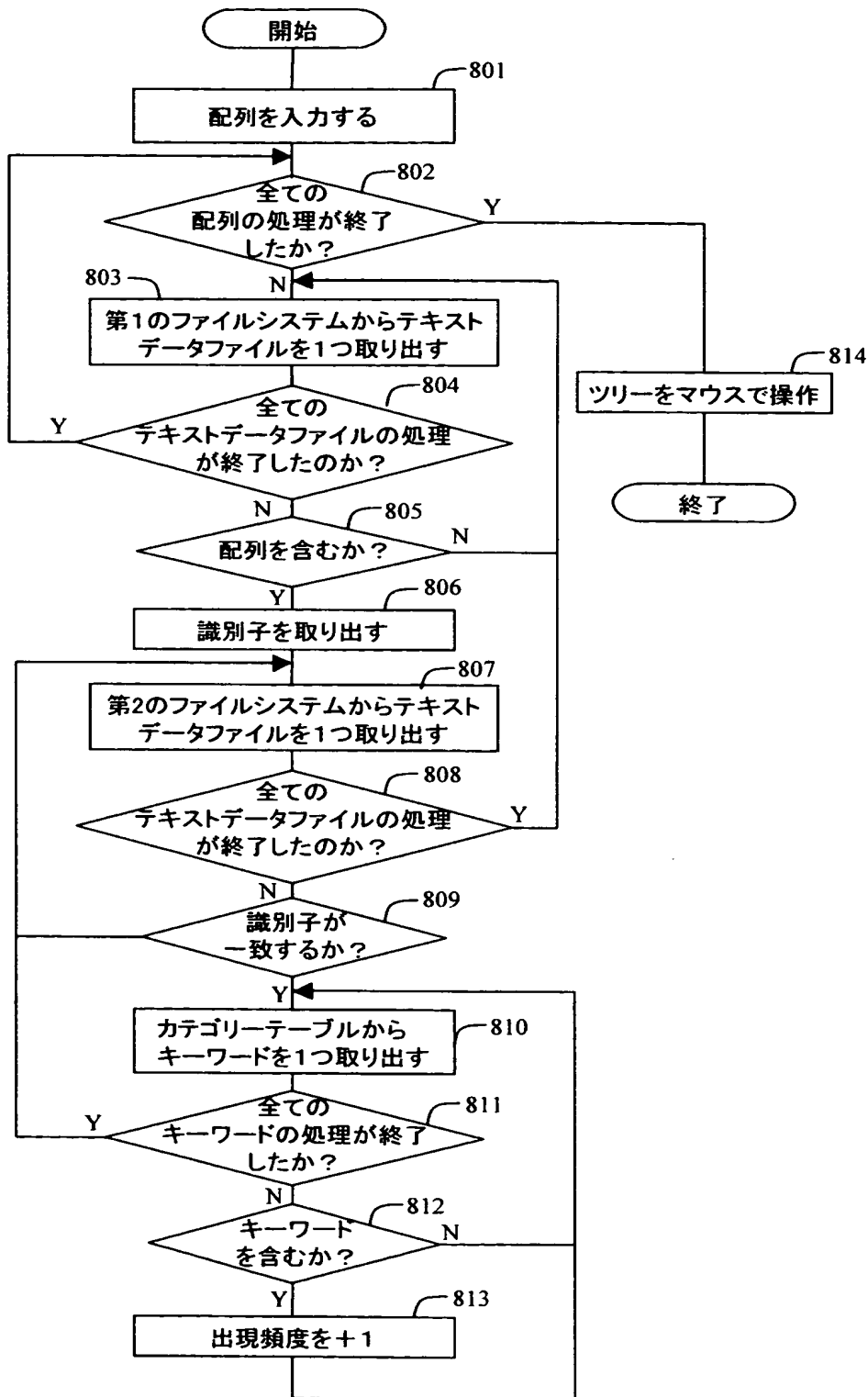
【図 6】

196	41	23
		18
	155	2
		153
1249	1233	39
		1194
	16	5
		11

【図 7】



【図 8】



【書類名】 要約書

【要約】

【課題】 塩基配列及びアミノ酸配列に関する情報を格納した第1のデータベースと文書データを格納した第2のデータベースとを用いてキーワードの出現頻度を算出する。

【解決手段】 キーワード頻度算出方法は、ユーザが入力した塩基配列又はアミノ酸配列に基づいて、第1のデータベースより第1のテキストデータを抽出する第1のテキストデータ抽出ステップと、上記抽出した第1のテキストデータよりそれに含まれる文書データを識別する識別子を抽出する識別子抽出ステップと、上記抽出した識別子に基づいて、第2のデータベースより第2のテキストデータを抽出する第2のテキストデータ抽出ステップと、キーワードテーブルよりキーワードを順次読み出し、上記第2のテキストデータにおける上記各キーワードの出現頻度を算出する出現頻度算出ステップと、を含む。

【選択図】 図1

特願 2 0 0 3 - 0 9 2 0 9 8

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 2 3 3 0 5 5 ]

1. 変更年月日

1 9 9 0 年 8 月 7 日

[変更理由]

新規登録

住 所

神奈川県横浜市中区尾上町 6 丁目 8 1 番地

氏 名

日立ソフトウェアエンジニアリング株式会社

2. 変更年月日

2 0 0 2 年 1 0 月 1 1 日

[変更理由]

住所変更

住 所

神奈川県横浜市鶴見区末広町一丁目 1 番 4 3

氏 名

日立ソフトウェアエンジニアリング株式会社